



Google

Associate-Data-Practitioner Exam

Google Cloud Associate Data Practitioner (ADP) Exam

Exam Latest Version: 6.2

DEMO Version

Full Version Features:

- 90 Days Free Updates
- 30 Days Money Back Guarantee
- Instant Download Once Purchased
- 24 Hours Live Chat Support

Full version is available at link below with affordable price.

<https://www.directcertify.com/google/associate-data-practitioner>

Question 1. (Single Select)

Your retail company wants to predict customer churn using historical purchase data stored in BigQuery. The dataset includes customer demographics, purchase history, and a label indicating whether the customer churned or not. You want to build a machine learning model to identify customers at risk of churning. You need to create and train a logistic regression model for predicting customer churn, using the customer_data table with the churned column as the target label. Which BigQuery ML query should you use?

A)

```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT *
FROM customer_data;
```

B)

```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT * EXCEPT(churned),
       churned AS label
FROM customer_data;
```

C)

```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT * EXCEPT(churned)
FROM customer_data;
```

D)

```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT churned as label
FROM customer_data;
```

- A: Option A
- B: Option B
- C: Option C
- D: Option D

Correct Answer: B

Explanation:

In BigQuery ML, when creating a logistic regression model to predict customer churn, the correct query should:

Exclude the target label column (in this case, `churned`) from the feature columns, as it is used for training and not as a feature input.

Rename the target label column to `label`, as BigQuery ML requires the target column to be named `label`.

The chosen query satisfies these requirements:

`SELECT * EXCEPT(churned), churned AS label`: Excludes `churned` from features and renames it to `label`.

The `OPTIONS(model_type='logistic_reg')` specifies that a logistic regression model is being trained.

This setup ensures the model is correctly trained using the features in the dataset while targeting the `churned` column for predictions.

Question 2. (Single Select)

You work for a healthcare company that has a large on-premises data system containing patient records with personally identifiable information (PII) such as names, addresses, and medical diagnoses. You need a standardized managed solution that de-identifies PII across all your data feeds prior to ingestion to Google Cloud. What should you do?

- A: Use Cloud Run functions to create a serverless data cleaning pipeline. Store the cleaned data in BigQuery.
- B: Use Cloud Data Fusion to transform the data. Store the cleaned data in BigQuery.
- C: Load the data into BigQuery, and inspect the data by using SQL queries. Use Dataflow to transform the data and remove any errors.
- D: Use Apache Beam to read the data and perform the necessary cleaning and transformation operations. Store the cleaned data in BigQuery.

Correct Answer: B

Explanation:

Using Cloud Data Fusion is the best solution for this scenario because:

Standardized managed solution: Cloud Data Fusion provides a visual interface for building data pipelines and includes prebuilt connectors and transformations for data cleaning and de-identification.

Compliance: It ensures sensitive data such as PII is de-identified prior to ingestion into Google Cloud, adhering to regulatory requirements for healthcare data.

Ease of use: Cloud Data Fusion is designed for transforming and preparing data, making it a managed and user-friendly tool for this purpose.

It's a fully managed, cloud-native data integration service for building ETL/ELT data pipelines visually.

It offers built-in transformations and connectors, including those suitable for data masking and de-identification.

It provides a standardized, visual interface, making it easier to create and manage data pipelines across various data sources.

It's designed for data integration and transformation, making it ideal for this scenario.

It helps to achieve a standardized managed solution.

Question 3. (Single Select)

You manage a large amount of data in Cloud Storage, including raw data, processed data, and backups. Your organization is subject to strict compliance regulations that mandate data immutability for specific data types. You want to use an efficient process to reduce storage costs while ensuring that your storage strategy meets retention requirements. What should you do?

- A: Configure lifecycle management rules to transition objects to appropriate storage classes based on access patterns. Set up Object Versioning for all objects to meet immutability requirements.
- B: Move objects to different storage classes based on their age and access patterns. Use Cloud Key Management Service (Cloud KMS) to encrypt specific objects with customer-managed encryption keys (CMEK) to meet immutability requirements.
- C: Create a Cloud Run function to periodically check object metadata, and move objects to the appropriate storage class based on age and access patterns. Use object holds to enforce immutability for specific objects.
- D: Use object holds to enforce immutability for specific objects, and configure lifecycle management rules to transition objects to appropriate storage classes based on age and access patterns.

Correct Answer: D

Explanation:

Using object holds and lifecycle management rules is the most efficient and compliant strategy for this scenario because:

Immutability: Object holds (temporary or event-based) ensure that objects cannot be deleted or overwritten, meeting strict compliance regulations for data immutability.

Cost efficiency: Lifecycle management rules automatically transition objects to more cost-effective storage classes based on their age and access patterns.

Compliance and automation: This approach ensures compliance with retention requirements while reducing manual effort, leveraging built-in Cloud Storage features.

Question 4. (Single Select)

You work for an ecommerce company that has a BigQuery dataset that contains customer purchase history, demographics, and website interactions. You need to build a machine learning (ML) model to predict which customers are most likely to make a purchase in the next month. You have limited engineering resources and need to minimize the ML expertise required for the solution. What should you do?

- A: Use BigQuery ML to create a logistic regression model for purchase prediction.
- B: Use Vertex AI Workbench to develop a custom model for purchase prediction.
- C: Use Colab Enterprise to develop a custom model for purchase prediction.
- D: Export the data to Cloud Storage, and use AutoML Tables to build a classification model for purchase prediction.

Correct Answer: A

Explanation:

Using BigQuery ML is the best solution in this case because:

Ease of use: BigQuery ML allows users to build machine learning models using SQL, which requires minimal ML expertise.

Integrated platform: Since the data already exists in BigQuery, there's no need to move it to another service, saving time and engineering resources.

Logistic regression: This is an appropriate model for binary classification tasks like predicting the likelihood of a customer making a purchase in the next month.

Question 5. (Single Select)

You are designing a pipeline to process data files that arrive in Cloud Storage by 3:00 am each day. Data processing is performed in stages, where the output of one stage becomes the input of the next. Each stage takes a long time to run. Occasionally a stage fails, and you have to address

the problem. You need to ensure that the final output is generated as quickly as possible. What

should you do?

- A: Design a Spark program that runs under Dataproc. Code the program to wait for user input when an error is detected. Rerun the last action after correcting any stage output data errors.
- B: Design the pipeline as a set of PTransforms in Dataflow. Restart the pipeline after correcting any stage output data errors.
- C: Design the workflow as a Cloud Workflow instance. Code the workflow to jump to a given stage based on an input parameter. Rerun the workflow after correcting any stage output data errors.
- D: Design the processing as a directed acyclic graph (DAG) in Cloud Composer. Clear the state of the failed task after correcting any stage output data errors.

Correct Answer: D

Explanation:

Using Cloud Composer to design the processing pipeline as a Directed Acyclic Graph (DAG) is the most suitable approach because:

Fault tolerance: Cloud Composer (based on Apache Airflow) allows for handling failures at specific stages. You can clear the state of a failed task and rerun it without reprocessing the entire pipeline.

Stage-based processing: DAGs are ideal for workflows with interdependent stages where the output of one stage serves as input to the next.

Efficiency: This approach minimizes downtime and ensures that only failed stages are rerun, leading to faster final output generation.



Full version is available at link below with affordable price.

<https://www.directcertify.com/google/associate-data-practitioner>

30% Discount Coupon Code: LimitedTime2025

The advertisement features a large yellow arrow pointing right, a man in a suit thinking with his hand on his chin, and three circular icons showing people working together. It includes text about a 30% discount coupon, money-back guarantees, and satisfied customers.

PDF

FREE TRIAL

* 100% MONEY BACK GUARANTEED

CERTIFICATION EXAMS STUDY GUIDES

* **Product Features**

- * 100% Success in the Final Exam
- * 90 Days Free Updates
- * Latest Exam Q/A
- * 24/7 Customer Support
- * Practice Exams

* **Free Demo for Practice Test & PDF**

50K Plus Satisfied Customers

VISA AMERICAN EXPRESS DISCOVER G Pay